

1/10

Sequence length 4052

CCAAGATTTAAAGCCCGCAAGTTTGTTCCTTGAGACCAGCGACTTTAGCTCCGATGCGGGAAGGAAAGCCGACCTCCGA
 TTTGGACATTTAAAGAGCTGGGCTTGAACCTCGTGAGTTTCGCTCTAAACTGCCCTTGAAATGAAGCTGGACTTGGAGG
 TGGCATGGAATATTCAATGGGAGAGCCGCGATGAGGCCGCCACCACGCTTCTGAAGGATGCCCGTGTGGAAGAATTT
 TGACGTGCCAGTGTCTCGTTCTACAGGGTGTTCATTCTTCCGCAATCTCAGAAAAATGGGACTAAAAGAACTATTT
 TGTAAATAAGAAGACTTCCATTTTAAATGACCAACATGTATTAAGATGGACACCTACTCTACGAAACACGAAGTTCTA
 TGGTCTCGAAGAAGCCCGTGCCTGTTAAAACTGATCCTAACTAAAAACAGACTTGAGTGGAT M R M L 4
 V S G R R V K K W Q L I I Q L F A T C F 12
 GTT AGT GGC AGA AGA GTC AAA AAA TGG CAG TTA ATT ATT CAG TTA TTT GCT ACT TGT TTT 24
 L A S L M F F W E P I D N E I V S H M K 44
 TTA GCG AGC CTC ATG TTT TTT TGG GAA CCA ATC GAT AAT CAC ATT GTG AGC CAT ATG AAG 132
 S Y S Y R Y L I N S Y D F V N D T L S L 64
 TCA TAT TCT TAC AGA TAC CTC ATA AAT AGC TAT GAC TTT GTG AAT GAT ACC CTG TCT CTT 192
 K H T S A G P R Y Q Y L I N H K E K C Q 84
 AAG CAC ACC TCA GCG GGG CCT CGC TAC CAA TAC TTG ATT AAC CAC AAG GAA AAG TGT CAA 252
 A Q D V L L L L P V K T A P E N Y D R R 104
 GCT CAA GAC GTC CTC CTT TTA CTG TTT GTA AAA ACT GCT CCT GAA AAC TAT GAT CGA CGT 312
 S G I R R T W G N E N Y V R S Q L N A N 124
 TCC GGA ATT AGA AGG ACG TGG GGC AAT GAA AAT TAT GTT CGG TCT CAG CTG AAT GCC AAC 372
 I K T L F A L G T P N P L E G E E L Q R 144
 ATC AAA ACT CTG TTT GCC TTA GGA ACT CCT AAT CCA CTG GAG GGA GAA GAA CTA CAA AGA 432
 K L A W E D Q R Y N D I I Q Q D F V D S 164
 AAA CTG GCT TGG GAA GAT CAA AGG TAC AAT GAT ATA ATT CAG CAA GAC TTT GTT GAT TCT 492
 F Y N L T L K L L M Q F S W A N T Y C P 184
 TTC TAC AAT CTT ACT CTG AAA TTA CTT ATG CAG TTC AGT TGG GCA AAT ACC TAT TGT CCA 552
 H A K F L M T A D D D I F I E M P N L I 204
 CAT GCC AAA TTT CTT ATG ACT GCT GAT GAT GAC ATA TTT ATT CAC ATG CCA AAT CTG ATT 612
 E Y L Q S L E Q I G V Q D F W I G E V H 224
 GAG TAC CTT CAA AGT TTA GAA CAA ATT GGT GTT CAA GAC TTT TGG ATT GGT CGT GTT CAT 672
 R G A P P I R D K S S K Y Y V S Y E M Y 244
 CGT GGT GCC CCT CCC ATT AGA GAT AAA AGC AGC AAA TAC TAC GTG TCC TAT GAA ATG TAC 732
 Q W P A Y P D Y T A G A A Y V I S G D V 264
 CAG TGG CCA GCT TAC CCT GAC TAC ACA GCC GGA GCT GCC TAT GTA ATC TCC GGT GAT GTA 792
 A A K V Y E A S Q T L N S S L Y I D D V 284
 GCT GCC AAA GTC TAT GAG GCA TCA CAG ACA CTA AAT TCA AGT CTT TAC ATA GAC GAT GTG 852

Fig. 1A

2/10

F M G L C A N K I G I V P Q D H V F F S	304
TTC ATG GGC CTC TGT GCC AAT AAA ATA GGG ATA GTA CCG CAG GAC CAT GTG TTT TTT TCT	912
G E G K T P Y H P C I Y E K M M T S H G	324
GGA GAG GGT AAA ACT CCT TAT CAT CCC TGC ATC TAT GAA AAA ATG ATG ACA TCT CAT GGA	972
E L E D L Q D L W K N A T D F K V K T I	344
CAC TTA GAA GAT CTC CAG GAC CTT TGG AAG AAT GCT ACA GAT CCT AAA GTA AAA ACC ATT	1032
S K G F F G Q I Y C R L M K I I L L C K	364
TCC AAA GGT TTT TTT GGT CAA ATA TAC TGC AGA TTA ATG AAG ATA ATT CTC CTT TGT AAA	1092
I S Y V D T Y P C R A A F I *	379
ATT AGC TAT GTG GAC ACA TAC CCT TGT AGG GCT GCG TTT ATC TAA	1137

TAGTACTGAATGTTGTATGTTTTCACTGTCAGTCAACCTGGATGAAAAAACCTTTAAATGTCGTCCTATACC
 CTAAGTAAATGAGGACGAAAGACAAATATTTTGAAAGCCTAGTCCATCAGAATGTTTCTTTGATTCTAGAAGCTGTTT
 AATATCACTTATCTACTTCATTGCCTAAGTTCATTTCAAAGAATTTGTATTTAGAAAAGGTTTATATTATTAGTGAAAA
 CAAAACTAAAGGGAAGTTCAAGTTCTCATGTAATGCCACATATATACTTGAGGTGTAGAGATGTTATTAAGAAGTTTGT
 ATGTTAGAATAATTGCTTTTGGAAAAATACCAATGAACGTACAGTACAACATTTCAAGGAAATGAATATATTGTTAGAC
 CAGGTAAGCAAGTTTATTTTGTAAAGAGCACTTGGTGGAGGTAGTAGGGGCAGGGAAGGTCAGCATAGGAGAGAAA
 GTTCATGAATCTGGTAAAAACAGTCTCTTGTCTTAAAGAGGAGATGTAGAAAAATGTGTACAATGTTATTATAAACAGAC
 AAATCACGCTTACCACATCCATGTAGCTACTGGTGTAGAGTCATTAAAAATACCTTTTTTTCATCTTTTTTCAAAGT
 TTAATGTGAACTTTGTAGAAAAGTGAATTAATGTTGCCCTAATACTTTATATGTTTTTAATGGATTTTTTTTAAAGTATTA
 GAAAAATGACACATAACACGGGCAGCTGGTTGCTCATAGGGTCTTCTCTAGGGAGAAACCAATTGTTAATCAAATAAGC
 TGATTTTAAATGACGTTTTCACTGGTTTTTAAATATTCAATATTGGTCTGTGTTTAAAGTTTGTATTGTAATGTAATTT
 ACATAGAGGAATATAATAATGGAGAGACTTCAAATGGAAAGACAGAACATTACAAGCCTAATGTCTCCATAATTTTATA
 AAATGAAATCTTAGTGTCTAAATCCTTGTACTGATTACTAAAAATTAACCCACTCTCTCCCAACAGGTCTTATAAACCA
 CAGCACTTTGTTCCAGTTTCAGAGTTTTAAATTGAGAGCATTAAACATCAAAGTTATAATATCTAAACCAATTTATTTT
 TCAATCAATAACTGTCAAGGTGATCTTTATTTCTAAATATTTCAAACCTGAAAAACAGAGTAAAAAAGTGATAGAAAAG
 TTGCCAGTTTGGGGTTAAAGCATTTTAAAGCTGCATGTTCCCTGTANTCAAAGAGATGTGTCTGAGATCTAATAGAGT
 AAGTTACATTTATTTTACAAAGCAGGATAAAAAATGTGGCTATAATACACACTACCTCCCTTCACTACAGAAAGAACTAG
 GTGGTGTCTACTGCTAGGGAGATTATATGAAGGCCAAAAATAATGACTTCAGCAAGAGTGACTGAACCTCACTCTAAGGCC
 TTTGACTGCAGAGGCACCTGTTAGGGAAAAATCAGATGTCTCATATAATAAGSTGATGTGGAAACACGCAAAACAAAAAC
 GAAAAAAGATTTCTCAGTATACACAACTGAATGATGATACTTACAATTTTGTAGCAGGTAGCTTTTAAATGTTTACAGAA
 ATTTAAATTTTCTATTTTGAAATTTGAGGCTGTTTACATTGCTTAGATAATTTAGAATTTTAACTAATGTCAA

Fig. 1B

3/10

ACTACAGTGTCAAACATTCTAGGTTGTAGTTACTTTCAGAGTAGATACAGGGTTTTAGATCATTACAGTTTAAAGTTTTC
TGACCAATTAAAAAACATAGAGAACAAAAGCATATTTGACCAAGCAACAAGCTTATAATTAATTTTATTAGTTGATT
GATTAATGATGTATTGCCTTTTGCCCATATATACCCGTGTATCTATACTTGGAAGTGTTAAGGTTGCCATTGTTGA
AAACATAAGTGTCTCTGGCCATCAAAGTGATCTTGTTCACGAGTGCTTTTGTGAAACAATTATTTATTTGCTGAAAG
AGCTCTTCTGAACTGTGTCCTTTTAATTTTGTCTTAGAATAGAATGGAACAAGTTTAAATTTCAAGGAAATATGAAGGC
ACTTCCTTTTTTTCTAAGAAGGAAGTTGCTAGATGATTCCTTCATCACACTTACTTAAAGTACTGAGAAGAGTATCTGT
AAATAAAAGGGTTCCAACCTTTAAAAAAGAAGGAAAAAACTTTTGGTGCTCCAGTGTAGGGCTATCTTTTAAAAAA
TGTCAACAAGGGGAAAAATAAACTATCAGCTTGGATGGTCACCTTGAATAGAAGATGCTTATACACAGTGTTATTGTTAAA
ATTTTTTTACCTTTTGGTTGGTTGCATCTTTTTCCATATTGTTAATTTATACCAAATGTAAATATTTGTATTAC
TTGAATTTTGCTCTTGTATGGCAAAATAATTAGTGAGTTAAAAAAAATCTATAGTTTCCAATAAACAACTGAAAAATT
AAAAAAA

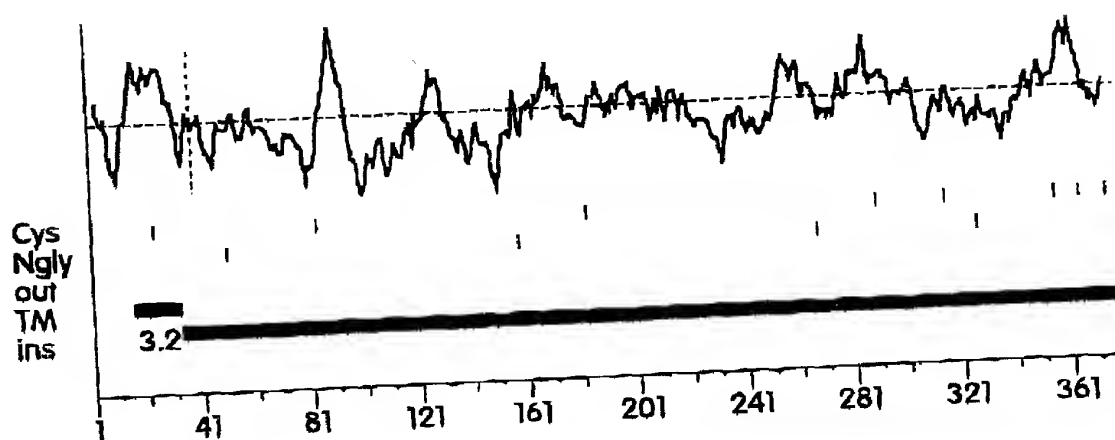
Fig. 1C

4/10

PFAM

no HMM hits

Galactosyl_T



>8797
MRMLVSGRRVKWQLIIQLFATCFLASLMFFWEPIDNHIVSHMKSYSYRYLINSYDFVND
TSLKETSAGPRYQYLINEKKEKCAQDVLILLFVKTAPEYDRRSGIRRTWGNENYVRSQ
LNANIKTLFALGTPNPLEGEELQKLAWEQRYNDIIQQDFVDSFYNTLTKLLMQFSWAN
TYCPHAKFLMTADDDIFIHMPNLIETLQSLQIGVQDFWIGRVHRGAPPTRDKSSKYVS
YEMYQWPAYPDYTAGAAYVISGDVAAKVYEASQTLNSSLYIDDVFMGLCANKIGIVPQDE
VFFSGEGKTPYHPCIEKMTSHGHLEDLQDLWKNATDPKVKTIKSGFFGQIYCRMKI
LLCKISYVDITYPCRAAPI

Fig. 2

5/10

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM

hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam5.4/Pfam

Sequence file: /prod/ddm/wspace/orfanal/oa-script.19955.seq

Query: 8797

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
Galactosyl_T	Galactosyltransferase	173.8	2.8e-48	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
Galactosyl_T	1/1	102	321	1	249	173.8	2.8e-48

Alignments of top-scoring domains:

Galactosyl_T: domain 1 of 1, from 102 to 321: score 173.8, E = 2.8e-48

```

      *--arEnaiRkTWmgnnsegvadgrikalPlvGl.sakgdqklklvme
      +rR iR+TW+n+n++++ ++ ik+lP +G++++++l++ + +
8797 102      DRRSGIRRTWGNENYVRSQNLNANIATLFAIGTNPLEGEELQKLAW 148

      KaktlyGDiiivDleDsYenLtlKtlitillygvskcpsakligKlDdDv
      E++ y Dii++D+ Dg++nLtlK l+ ++++++cp+ak+ + DdD+
8797 149      EDQ--RYNDIIQQDFVDSFYNLTLKLLMQFSWANTYCPHAKFLMTADDDI 196

      fvnpdkLlslLereniridpsessfyGylikegepvrkkskrdWVvppt
      ft +L++L++ i +++++ G++++ +p+r k sk Yv++
8797 197      FIEMPNLIEYLQSL-EQIGVQDFWI-GRVERGAPPIDKSSK--YVVSYE 242

      eYpcsrNgnkYpYvsGpFYllbrdAAplIlkaskhrLr.flkiedVlit
      Y + YP Y +G Y++s+d+A +++++ + ++ l i+DV++
8797 243      MYQWPA----YPDYTAGAAYVISGDVAAKVYEASQTL-NsSLYIDDVFM- 286

      GilaedlgIsrinlprlsistnlfrfhhsqkdndgcdvfawbtahkndpe
      G +a++gI +++ +f++ +++ h++ +e
8797 287      GLCANKIGIVPQDH-----VFFSGEGKTPY-----HPCIYE 317

      ylif<-+
      ++ +
8797 318      KMMT 321

```

Fig. 3

6/10

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
15	32	out-->ins	3.2

>8797

MRMLVSGRRVKKWQLIIQLFATCFLASLMFFWEPIDNHIIVSHMKSSYSRYLINSYDFVND
TLSLKHTSAGPRYQYLINHKEKCQAQDVLLLLFVRTAPENYDRRSGIRRTWGNENYVRSQ
LNANIKTLFALGTPNPLEGEELQORKLAWEDQRYNDIIQQDFVDSFYNTLTKLLMQFSWAN
TYCPHAKFLMTADDDIFIHMPNLI EYLOSLEQIGVQDFWIGRVHRGAPPIRDKSSKYYVS
YEMYQWPAYPDYTAGAAYVISGDVAARKVYEASQTLNSSLYIDDVFMGLCANKIGIVQDH
VFFSGEGKTPYHPCYKMTSRGHLEDLQDLWRNATDPKVKTIKGFQIYCRIMKII
LLCKISYVDTPCRAAFI

Fig. 4

7/10

Phase 1.3.3 Expression of 8797 w/β2

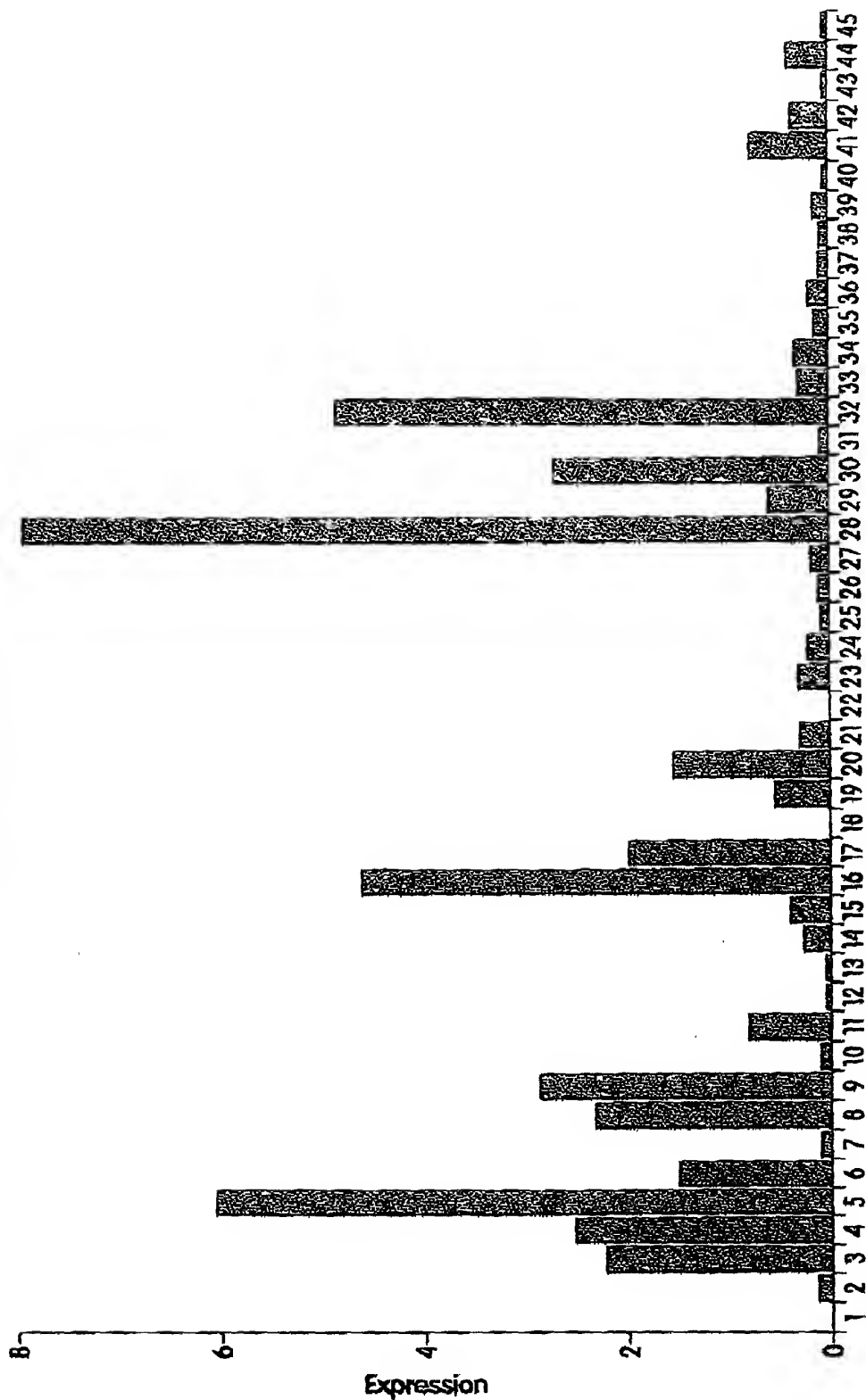


Fig. 5

8/10

8797 Expression in Oncology Phase II Plate

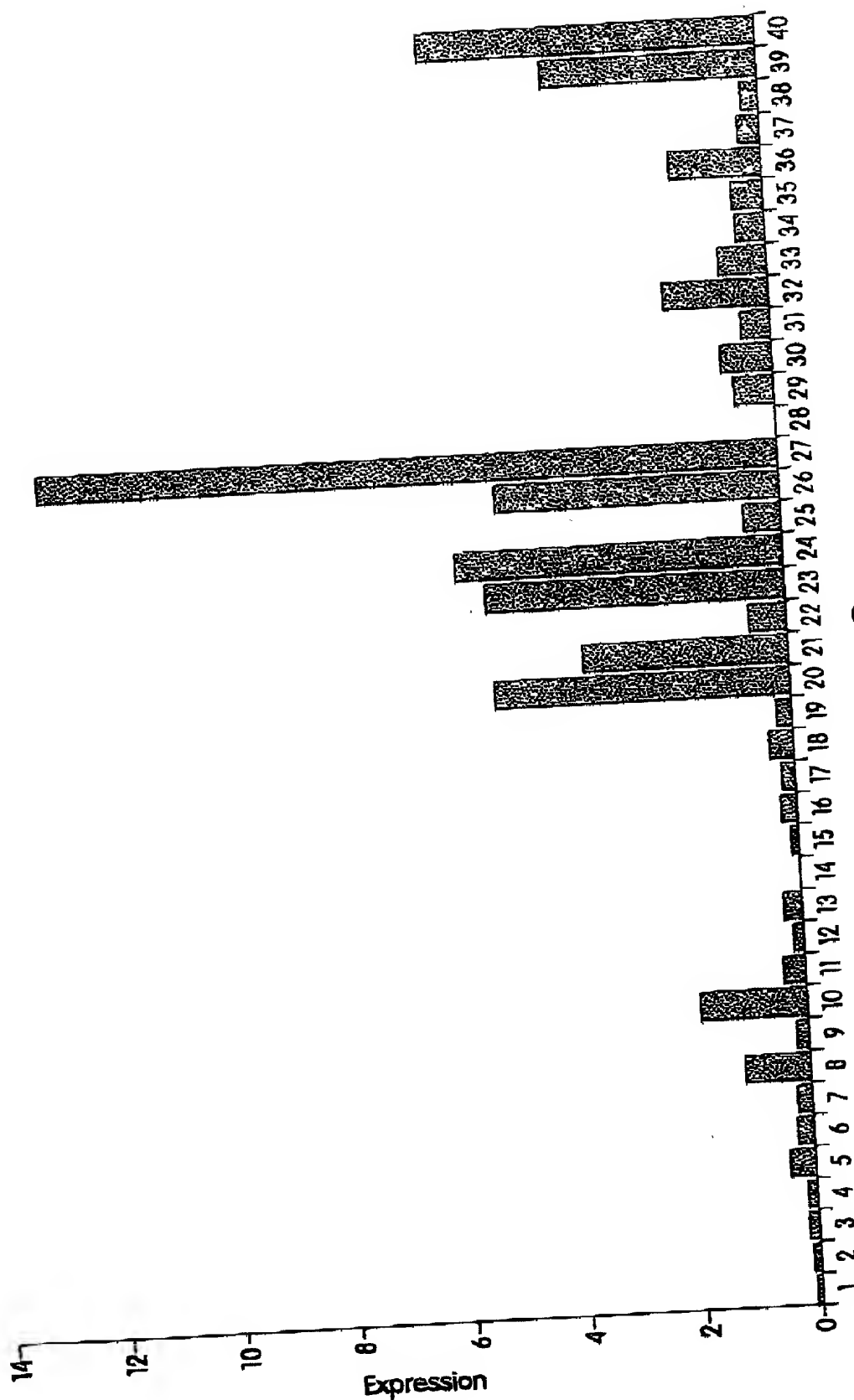


Fig. 6

8797 Expression in Lung Model Panel

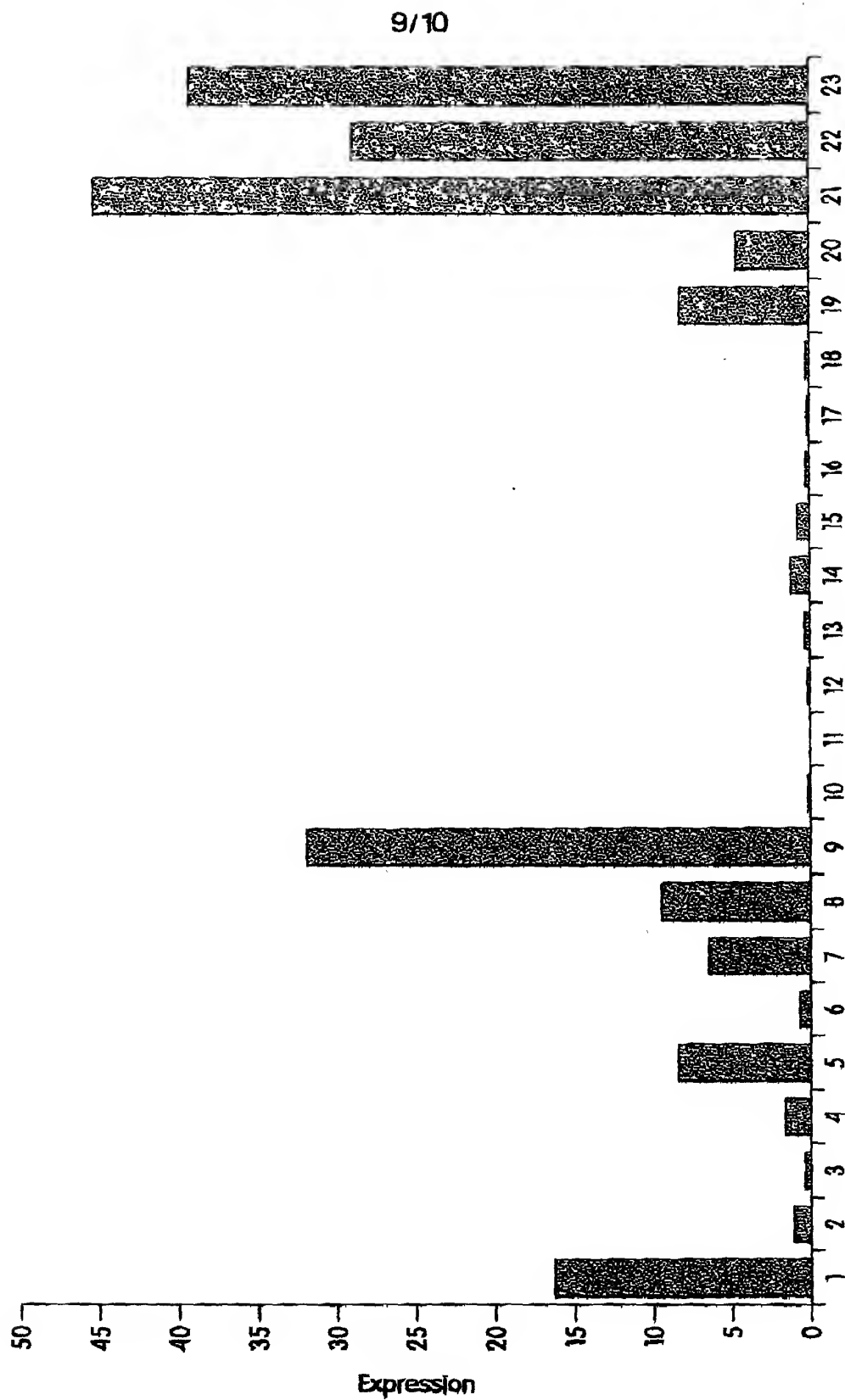


Fig. 7

10/10

8797 Expression in the Breast Models Panel

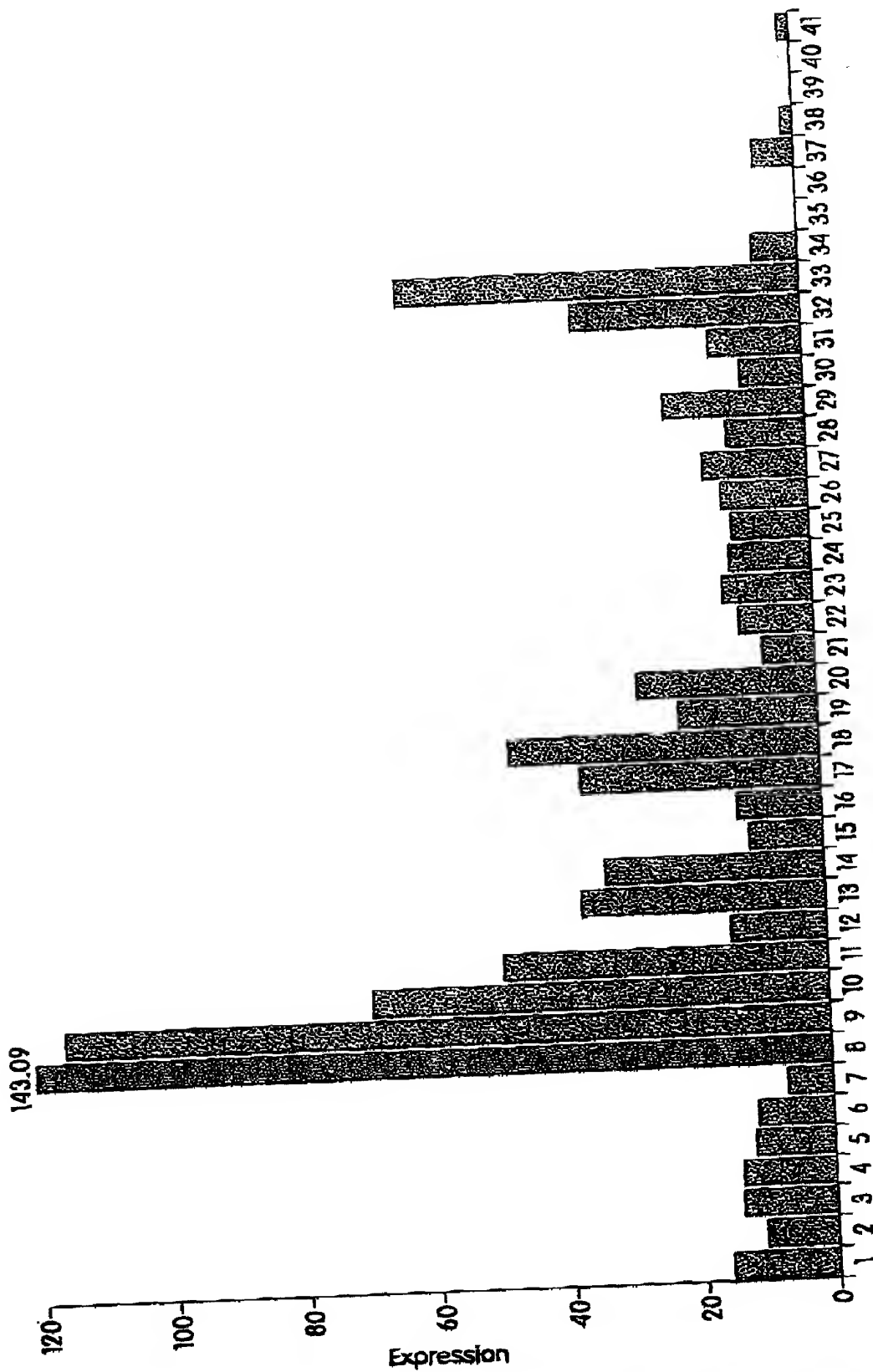


Fig. 8